MOLECULAR AND CELLULAR BIOLOGY, Oct. 1992, p. 4251–4261
0270-7306/92/104251-11$02.00/0

Vol. 12, No. 10

Cloning of GT Box-Binding Proteins: A Novel Sp1 Multigene Family Regulating T-Cell Receptor Gene Expression

CHRIS KINGSLEY AND ASTAR WINOTO*

Department of Molecular and Cell Biology, Division of Immunology, and Cancer Research Laboratory, University of California, Berkeley, Berkeley, California 94720

Received 5 May 1992/Returned for modification 26 June 1992/Accepted 7 July 1992

Analysis of a T-cell antigen receptor (TCR) α promoter from a variable gene segment (V) revealed a critical GT box element which is also found in upstream regions of several Vα genes, TCR enhancer, and regulatory elements of other genes. This element is necessary for TCR gene expression and binds several proteins. These GT box-binding proteins were identified as members of a novel Sp1 multigene family. Two of them, which we term Sp2 and Sp3, were cloned. Sp2 and Sp3 contain zinc fingers and transactivation domains similar to those of Sp1. Like Sp1, Sp2 and Sp3 are expressed ubiquitously, and their in vitro-translated products bind to the GT box in TCR Vα promoters. Sp3, in particular, also binds to the Sp1 consensus sequence GC box and has binding activity similar to that of Sp1. As the GT box has also previously been shown to play a role in gene regulation of other genes, these newly isolated Sp2 and Sp3 proteins might regulate expression not only of the TCR gene but of other genes as well.

T lymphocytes can be subdivided into two populations of αβ and γδ T cells on the basis of their T-cell receptor (TCR) expression and function. The T-cell receptor is a heterodimeric protein composed of either αβ or γδ subunits with which T cells can recognize foreign antigens. Each of the genes encoding all four TCR subunits is composed of several gene segments (V [variable], J [joining], D [diversity], and C [constant]) which rearrange during development to form VDJ or VJ segments. Subsequent transcription and RNA splicing then form a mature mRNA, which, if containing a proper open reading frame, gives rise to a functional polypeptide containing VDJC or VJC regions (for reviews, see references 5 and 27). The process of V-D-J gene rearrangements is developmentally and tissue specifically regulated. The TCR α and δ genes are particularly interesting because the δ locus is located within the TCR α-gene segments. In T cells committed to the αβ lineage, the α-specific V-gene segments are rearranged to the α locus, whereas in cells committed to the γδ lineage, the δ-specific V-gene segments are rearranged to their respective D/Jδ gene segments. Some V gene segments, however, can rearrange to either the α or δ locus (these gene segments are also called Vα).

Studies using transgenic mice and cell lines suggest that specific gene segment transcription correlates with differential accessibility of each locus to the recombinase during development (8, 7). Production of sterile transcripts from unrearranged gene segments, in particular, has been found to precede gene rearrangement. It is hypothesized that sterile transcription is either the cause or the consequence of differential opening of chromatin structure and hence accessibility of the gene segment to the recombination enzymes (references 7 and 28 and references therein). Since each of the V-gene segments contains its own promoter, it is possible that the specificity of sterile transcription from these V promoters could contribute to differential accessibility to recombinases during T-cell development, which could result in specific gene segment usage in αβ and/or γδ T cells (35). Therefore, the study of TCR α promoters can serve the dual purpose of analyzing TCR gene expression as well as regulation of rearrangement.

We have focused on the well-characterized Vα11.1 gene segment, which is the gene segment predominantly used in T-cell response to pigeon cytochrome *c* (9, 38). We have analyzed the critical elements for the Vα11.1 promoter and found one element which we called the GT box. This GT box is also present in most other available Vα upstream sequences (presumably promoters) as well as in the TCR α enhancer. Homologous regulatory elements (AC boxes) were previously defined in the β-globin enhancers, promoters, and locus-controlling region (LCR) as well as in the bovine papillomavirus (BPV) promoters and simian virus (SV40) enhancer. This novel element for T-cell gene expression was shown to bind ubiquitously expressed proteins, one of which was identified as the transcription factor Sp1 (18). Although Sp1 is one of the earliest-characterized transcription factors which also binds to a consensus GC box sequence, we have made the assumption that additional GT box-binding proteins contain sequence homology to the Sp1 DNA-binding domain. We have isolated two novel Sp1-related genes, which we call Sp2 and Sp3. Sequence analysis of the Sp2 and Sp3 cDNA clones predicts that they encode proteins with several transactivation domains and a Zn finger DNA-binding domain with extensive homology to Sp1. Their in vitro-produced proteins also bind to the GT box. Hence, Sp1, Sp2, and Sp3 constitute a novel Sp1 multigene family with similar characteristics which can presumably regulate a wide variety of genes in different tissues. Furthermore, the existence of proteins similar to Sp1 indicates that gene regulation by Sp1 is more complicated than previously assumed.

MATERIALS AND METHODS

Plasmid construction and transfection. The Vα11.1 promoter region was obtained by using a limited genomic library from the T-cell hybridoma AN6.2 reactive for cytochrome *c* as described previously (38).

Plasmid pCATMoEnT was constructed in three steps.

* Corresponding author.

First, a 435-bp *Cla*I-to-*Xba*I fragment of the Moloney virus enhancer with the *Cla*I site converted to *Sal*I was subcloned into plasmid pUC18 (pUCMoEn). Second, the Moloney enhancer from pUC-MoEn was excised with *Sal*I and *Sma*I and inserted into *Xho*I-*Eco*RV sites of plasmid pCAT3'L, generating pCATMoEn (pCAT3'L is a derivative of pUC13 with the chloramphenicol acetyltransferase [CAT] gene inserted between the *Sma*I and *Eco*RI sites and a polylinker containing *Xho*I-*Eco*RV-*Bgl*II-*Cla*I sites at the *Nde*I site 3' of the CAT gene). Third, a 237-bp *Bcl*I-*Bam*HI fragment of a transcriptional terminator from the SV40 poly(A) site was subcloned into the *Bam*HI site of plasmid pSP72 (Promega). The *Pvu*II site of the resulting plasmid was converted to a *Cla*I site, and the terminator was excised with *Cla*I. The terminator *Cla*I fragment was then cloned into the *Acc*I site of pCATMoEn, generating plasmid pCATMoEnT.

Plasmid pTCR1700 was constructed in two steps. A 1.7-kb *Hin*cII-partial *Hae*III fragment of the Vα11.1 promoter sequences was subcloned into the *Hin*cII site of plasmid pUC18 to generate construct 18pTCR11; this plasmid was then digested with *Hin*dIII-*Sma*I to liberate the Vα11.1 promoter sequence, which was ligated into a 5.2-kb *Hin*dIII-*Sma*I fragment of plasmid J21 (36). pTCR1700MoEn was made by inserting the *Sal*I-*Sma*I Moloney enhancer fragment of pUCMoEn into *Xho*I-*Eco*RV sites of plasmid pTCR1700.

The following plasmids were made in pCATMoEnT: pTCR250MoEnT, pTCR150MoEnT, pVα2C470MoEnT, and pVα2C240MoEnT. For pTCR250MoEnT and pTCR150Mo EnT, a blunt-ended *Bst*NI-*Sma*I fragment and a *Pvu*II-*Sma*I fragment of pTCR350MoEn (=pTCR650MoEn) containing the Vα11.1 promoter sequences were used (pTCR350MoEn/pTCR650MoEn is similar to pTCR1700MoEn, with the 5' end extending to *Pst*I site only). For the Vα2C promoter (10), a 470-bp *Sty*I fragment and a 240-bp *Bgl*II-*Sty*I fragment from the Vα2C promoter region were blunt ended with Klenow enzyme and cloned into the *Sma*I site of plasmid pCATMoEnT to generate pVα2C470MoEnT and pVα2 C240MoEnT, respectively.

The start of Vα11.1 transcription was estimated by RNase protection. For this purpose, a *Bst*NI-*Xba*I fragment of plasmid pTCR1700 was cloned into the *Sma*I site plasmid pSP72 (pTCR1700 was made by inserting *Sma*I-*Hin*dIII fragment of 18pTCR11 into the corresponding sites of pCAT). The plasmid was linearized with *Eco*RI, and an SP6 transcript was generated. As a marker, a Maxam-Gilbert G sequencing reaction was performed on a *Pvu*II-*Xba*I fragment of plasmid pTCR1700. Plasmids J21 and J21MoEn containing a minimal c-*fos* promoter have been described before (36). Transfection into various cell lines and subsequent assays for CAT activity were performed as described elsewhere (34).

Heteroduplex mutagenesis. Heteroduplex oligonucleotide-mediated mutagenesis was performed in double-stranded plasmid DNA as described previously (22). The GT box sequence GCAGAGGTGGGTGGAGTTTCG in plasmid pTCR250MoEnT was changed to GCAGAGGTTTTTGGAG TTTCG to generate plasmids mut-1 and mut-2. Mutant plasmids were screened with a mutant oligonucleotide at high stringency and confirmed by DNA sequencing (mut-1 is identical to mut-2). For gel mobility shift assays of GT box-binding protein, oligonucleotides VA1 (TCGAGAG GTGGGTGGAGTTTCGCG) and VA2 (TCGACGCGAAA CTCCACCCACCTC) were annealed and labeled with Klenow enzyme. For the mutant GT box oligonucleotides, 2 bp of the GGTGGGTGG sequences in VA1 were changed to GGTTTGTGG (mVA2). The corresponding sequences in the VA2 oligonucleotide were also changed (mVA3). The two mutant oligonucleotides were then annealed and used in gel mobility shift experiments.

Gel shift analysis and methylation interference. Gel shift assays were performed as described previously (30, 33), using 0.5× Tris-borate buffer. Methylation interference analyses were performed as described elsewhere (29, 30). For purified Sp1 protein, binding to the GC box was facilitated by the inclusion of 0.1% Nonidet P-40 and 0.1 mg of bovine serum albumin (BSA) per ml in the binding reaction mixtures (14). Nonidet P-40 and BSA were not needed, however, for Sp1 produced by in vitro translation. Purified human Sp1 protein from HeLa cells infected with recombinant vaccinia virus containing a full-length Sp1 cDNA was purchased from Promega. For the Sp1 consensus probe (GC box), the two complementary oligonucleotides corresponding to the sequence ATTCGATCGGGGCGGGGCGAGC (Promega) were kinase treated and used in the gel mobility shift assay. For the GT box binding reaction, the buffer consisted of 10 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES)-KOH (pH 7.9), 50 mM NaCl, 5 mM Tris-HCl (pH 7.5), 15 mM EDTA, 1 mM dithiothreitol, and 5% glycerol. For competition studies, the following subclones were generated for the different Vα promoters: Vα11.1 (*Bst*NI-*Pvu*II fragment into *Pvu*II of pSP72), Vα2C (*Eco*RI-*Bam*HI fragment into the corresponding sites in pSP72) (pSP72Vα2C), VαBDFL1 (*Bam*HI-*Kpn*I fragment into *Bam*HI-*Kpn*I sites of pSP72) (pSP72-BDFLI), and Vα5H (*Hin*dIII fragment into *Hin*dIII site of pUC8). For nonspecific competition, oligonucleotides 12RSS-A (GATCCCACGTGCTCCAGGGCTG AACAAAAACCGA) and 12RSS-B (GATCTCGGTTTTTG TTCAGCCCTGGAGCACTGTGG) were used.

Cloning, in vitro transcription and translation, and gel shift. For cloning of Sp2 and Sp3, a human T-cell library (HUT78 plus phorbol myristate acetate) was screened at low stringency, using the *Nco*I fragment containing the Sp1 Zn finger. Washes were performed at 50°C in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) for low-stringency conditions and at 65°C in 0.1× SSC for high-stringency conditions. Two clones, S1 and S13, corresponding to Sp2 and Sp3, respectively, were obtained. Both were subcloned into pSP72 vector (Promega). To obtain a full-length Sp2 clone, another T-cell cDNA library from Molt13 (constructed in λgt11) was screened. A clone containing a 2.7-kb insert, S1/7, was obtained and subcloned into pBluescript II-KS(−) vector (Stratagene). In vitro transcription and translation were performed as suggested by the manufacturer (Promega). Northern (RNA) blotting was performed by using S1 and S13 inserts as probes as described previously (24).

Nucleotide sequence accession numbers. Sequences reported in this paper have been submitted to the GenBank and EMBL data bases and have been assigned accession numbers as follows: Sp2 and Sp3 DNAs, M97190 and M97191, respectively (GenBank); Vα11.1 and Vα5H promoters, X62486 and X62487, respectively (EMBL).

RESULTS

Analysis of the TCR Vα promoters. To define the TCR α-promoter regulatory element(s), we analyzed the Vα11.1 promoter in a transient transfection experiment using the reporter CAT gene. A 1.7-kb upstream region of the Vα11.1 promoter driving the CAT gene was inactive (pTCR1700) in all cell lines examined (Jurkat, YAC-1, BJA-B, and S194;

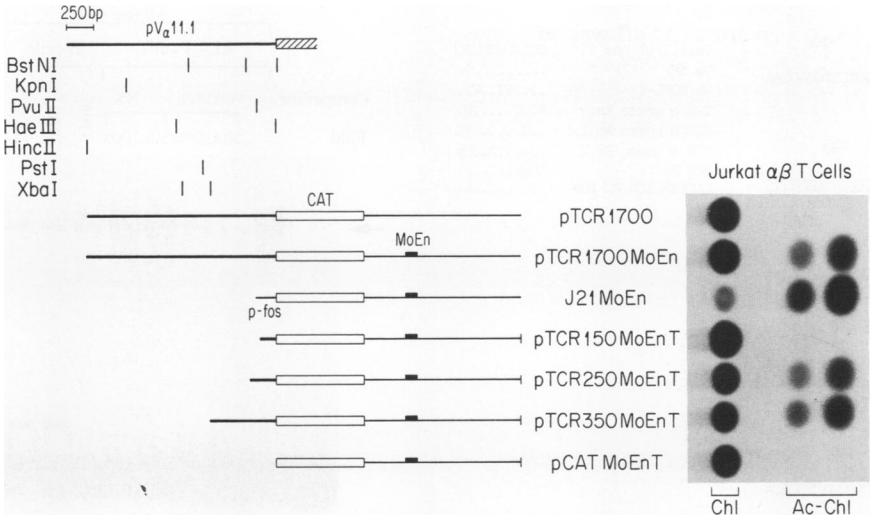

FIG. 1. Deletional analysis of the T-cell-specific TCR Vα11.1 promoter. Various DNA fragments corresponding to the Vα11.1 promoter region (pVα11.1) were fused to the CAT gene and tested for promoter activity in the presence of a strong Moloney virus enhancer (MoEn). Promoter activity was tested in T cells (Jurkat and YAC-1). For short promoter sequences, a transcription terminator was placed upstream to avoid any possible read-through due to the presence of a strong Moloney virus enhancer (pTCR150MoEnT, pTCR250MoEnT, and pTCR350MoEnT). A negative control with a promoterless plasmid (pCATMoEnT) and a positive control with the c-*fos* promoter sequences (J21MoEn) were included in the experiment. Chl, [^{14}C]chloramphenicol; Ac-Chl, acetylated [^{14}C]chloramphenicol.

Fig. 1 and data not shown). This result is consistent with the notion that the TCR promoter alone is very weak and is activated by the enhancer located near the constant region after rearrangement has taken place. When a heterologous enhancer from the Moloney murine virus long terminal repeat was added to the 3′ end of the constructs (pTCR1700MoEn and pTCR350MoEnT), however, CAT activity could be detected in all T-cell lines examined (Jurkat, YAC-1, and EL4; Fig. 1 and unpublished result). For all the experiments, a c-*fos* promoter and Moloney virus enhancer construct (J21MoEn) which works in all cell lines was used as a positive control, and a promoterless Moloney virus enhancer construct was used as a negative control (pCATMoEnT). The minimal functional promoter elements (pTCR250 MoEnT in Fig. 1) were found within 130 bp 5′ of the major RNA initiation site, as determined by an RNase protection experiment (Fig. 2A). An essential promoter element was further localized to an 88-bp fragment (−130 to −42), since deletion of the corresponding DNA resulted in a completely inactive promoter (Fig. 1, pTCR150MoEnT).

Sequence analysis of this minimal Vα promoter region did not reveal a TATA box, GC-rich region, or initiator consensus sequence (YYCAYYYYY [31]). A purine-rich region was found at positions −36 to −31, but this region is insufficient for promoter activity (pTCR150MoEnT). Scanning of the essential promoter region from −130 to −42 revealed a conserved region among several other Vα upstream sequences (VαBDFL-1 and Vα5H; only two upstream Vα sequences [mouse Vα BDFL-16, and mouse Vα2C 10] are available in GenBank. The Vα5H upstream sequences were determined in this laboratory). This DNA sequence is similar to the AC box found in the β-globin LCRs (hypersensitive site 2 [HSS2], HSS1, HSS3, and HSS4), α-globin and γ-globin promoters, β-globin enhancer (consensus sequence GNNGGGTGGRGYSN [R = purine, Y = pyrimidine, S = C or G, and N = any nucleotide [26]), and the BPV regulatory elements (23) (Fig. 2B). In addition, the element is found in the TCR α enhancer (≅100 bp 3′ of the GATA3 site in the mouse genome and ≅60 and 100 bp 3′ of the GATA3 site in the human genome; Fig. 2B).

GT box-binding proteins. To identify the protein(s) regulating the activity of the TCR α promoter, we performed DNase I footprint analysis with the minimal promoter region. No DNase I-protected regions, however, were found (33a). As an alternate approach, a gel shift assay using the minimal 88-bp *Bst*NI-*Pvu*II fragment as a probe was performed. One major retarded protein-DNA band was found in a variety of nuclear extracts from human and mouse lymphoid cell lines (EL4, Jurkat, Molt13, Peer T cell, S194, 38B9, 22D6, and BJA-B B cell) and nonlymphoid cell lines (L cells and HeLa; arrow in Fig. 3). The protein-DNA interaction is specific, as the complex could be competed away with 50- and 100-fold excesses of specific probe (Fig. 3, lanes 3 and 4) but not with an equivalent amount of nonspecific DNA (lanes 5 and 6). Additional complexes were found in nuclear extracts of human origin; these complexes, however, are probably nonspecific protein-DNA complexes, as they cannot be competed for with a 400-fold excess of a specific DNA fragment (19a). The protein-binding site was identified by methylation interference and localized to the conserved GT box spanning positions −48 to −64 of the Vα11.1 promoter (Fig. 4). The same methylation interference pattern was seen with use of nuclear extracts from T cells of either human or mouse origin (Jurkat or EL4, respectively). Hence, we have identified a protein complex which binds to the conserved GT box of the TCR α promoters.

As discussed above, homology to the GT box was found in human and mouse TCR α enhancers and two other Vα upstream sequences at positions similar to that of the GT box in the Vα11.1 promoter. This homology is significant because competition studies performed in gel shift experiments indicated that the GT box DNA-protein complex can be competed for with an excess of DNA fragments from the TCR α enhancer, Vα BDFL-1 and Vα5H, albeit with different degrees of efficiency (data not shown). No strong homol-

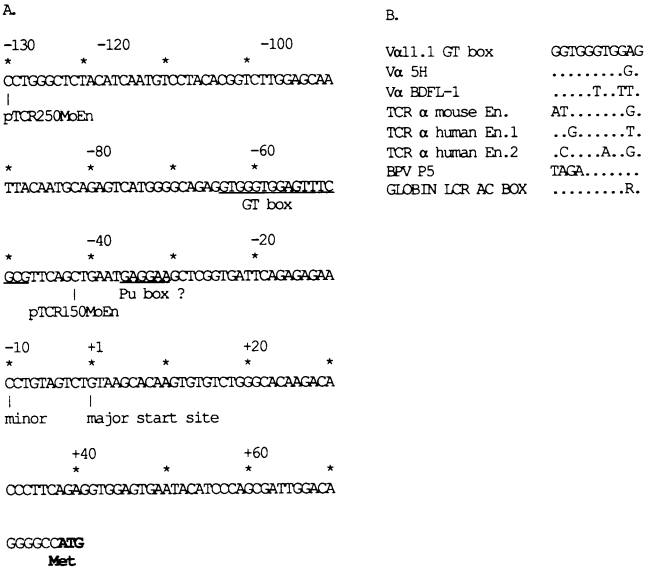

FIG. 2. DNA sequences of the Vα11.1 minimal promoter and comparison of GT boxes. (A) DNA sequence of the minimal functional Vα11.1 promoter, determined by the dideoxy sequencing technique. The boundary of pTCR150MoEn is indicated. The locations of the GT box (see below) and a possible Pu box are underlined. Two RNA start sites are indicated, with the major start site labeled +1. RNase protection experiment was performed to define the RNA start site of the Vα11.1 promoter. RNA from hybridoma BC15.1, which contains a Vα11.1 rearranged TCR α gene, was used to hybridize to an SP6 in vitro-transcribed Vα11.1 RNA. Specificity control was performed with HeLa RNA. The G ladder of the Vα11.1 promoter region from plasmid pTRC1.7 (see Materials and Methods) was used as a marker. (B) Homology of the GT box element to other Vα promoters, the TCR α enhancer, and other regulatory elements. Sequences of the Vα11.1 GT box (≅140 bp from AUG) were compared with upstream DNA sequences from Vα5H (≅160 bp from AUG [37]), VαBDFL-1 (≅100 bp from AUG [6]), mouse TCR α enhancer (≅100 bp from the GATA3-*Bgl*II site; sequences 3′ of the mouse TCR α enhancer can be found in GenBank under accession number M64239 [submitted by L. Hood]), human TCR α enhancer (homology 1 and homology 2 are ≅60 and 100 bp, respectively, from the GATA3 site [13]), BPV P5 promoter region (23), and β-globin LCR AC box (26). Dots indicate identity.

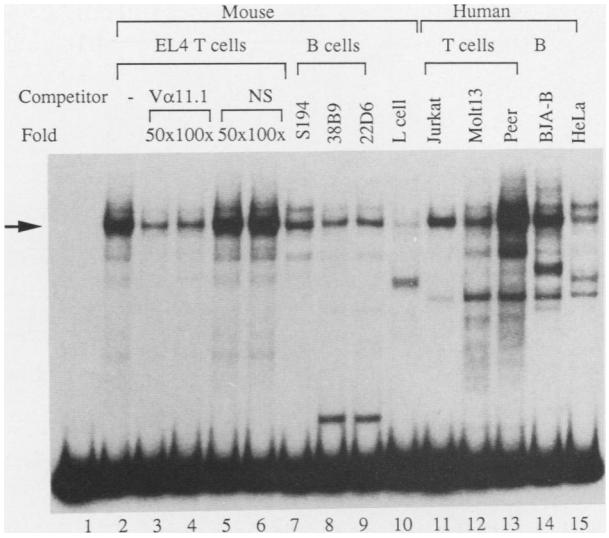

FIG. 3. Gel shift analysis of the minimal Vα11.1 promoter fragment. Gel mobility shift analysis was performed by using a labeled *Bst*NI-*Pvu*II fragment of the Vα11.1 promoter region. Specificity of the DNA-protein interaction was tested with 50- and 100-fold molar excesses of unlabeled *Bst*NI-*Pvu*II fragment and control oligonucleotides containing heptamer-nonamer sequences of the immunoglobulin signal for rearrangement (12RSS). Nuclear extracts from the indicated T cells, B cells, pre-B cells, and nonlymphoid cells were used. The GT box-binding protein–DNA complex is indicated by an arrow.

ogy or competition, however, could be found with the third Vα 2C upstream sequences (Vα3 [10]), although the region contains a functional Vα promoter (pVα2C240MoEnT and pVα2C470MoEnT, containing 240 and 470 bp of the Vα2C promoter region, respectively; Table 1). Hence, the GT box is conserved only in a subset of Vα promoters, consistent with the notion that Vα promoters are heterogeneous.

The functional significance of the GT box was assessed by site-directed DNA mutagenesis. Three nucleotides of the GT box were changed in the Vα11.1 promoter-Moloney enhancer construct (pTCR250MoEnT) to generate mutant Vα promoter-Moloney enhancer constructs (mut-1 and mut-2 are two independent isolates with identical base pair mutations). The mutation abolished the protein-binding site, as judged by a gel mobility shift assay of the mutant GT box oligonucleotide (data not shown). Transient transfection experiments of the mutated constructs showed that mutation at the GT box eliminated the Vα promoter activity altogether (Table 1). Thus, the GT box binding site is an essential element for TCR α-gene expression.

At least two proteins bind to the GT box, one of which is Sp1. The GT box of the Vα11.1 promoter is closely related to the P4 and P5 regulatory elements of BPV, there being a 10-nucleotide identity between them. The BPV element has been shown recently to bind to the ubiquitously expressed transcription factor Sp1 (18), although the binding site is different from the Sp1 consensus sequence (23, 32). To probe the relationship between Sp1 and the GT box-binding proteins, we performed gel shift analysis under several conditions, using purified Sp1 protein and crude nuclear extracts containing GT box-binding proteins. Purified Sp1 protein can indeed bind to the GT box. However, the methylation interference pattern with use of purified Sp1 and the Vα promoter fragment is different from that of nuclear extracts (Fig. 4). Upon further analysis using a small GT box oligonucleotide as a probe, two DNA-protein complexes could be resolved in all cell lines tested (Fig. 5). The relationship between Sp1 and the GT box proteins from the nuclear extracts was further investigated by using a monoclonal antibody against human Sp1 generated with peptide sequences immediately outside the Sp1 Zn finger (14). The monoclonal antibody specific for human Sp1 recognized and blocked the binding activity of one of the GT box proteins from a wide variety of human nuclear extracts (T cells [Jurkat and Molt13], B cells [BJA-B], and nonlymphoid cells [HeLa]; Fig. 5). The effect of the antibody was specific for human Sp1, as addition of the antibody did not affect the binding activity of GT box proteins from a mouse nuclear extract (EL4). Hence, the GT box element in the Vα promoter binds to at least two different proteins, one of which is most likely Sp1.

Cloning of Sp2 and Sp3 with sequence homology to Sp1: evidence for an Sp1 multigene family. Since Sp1 can bind to the GT box, it was assumed that some other GT box-binding protein(s) would be homologous to Sp1 in its DNA-binding

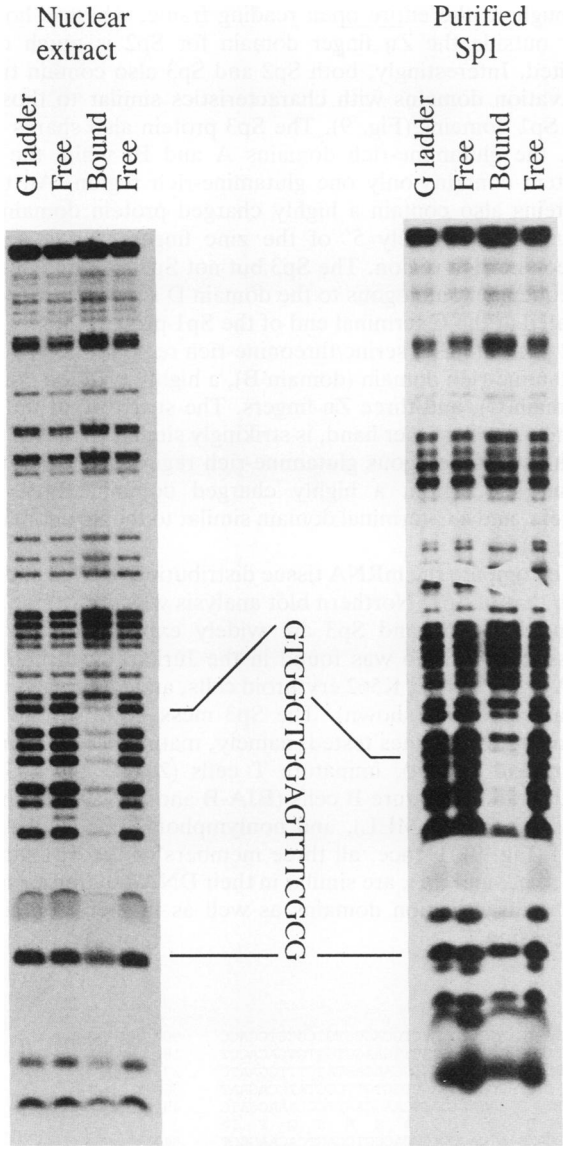

FIG. 4. Methylation interference of the promoter-binding protein. Methylation interference of the GT box-binding protein was done with the *Bst*NI-*Pvu*II fragment as a probe. Interference patterns of both nuclear extracts and purified Sp1 protein on the coding strand are shown. Both coding and noncoding strands were tested for interference with binding of the GT box-binding protein from EL4 and Jurkat T cells, but only the coding strand showed any interference. DNA sequences bound to the GT box-binding proteins from nuclear extract and purified Sp1 protein are indicated.

TABLE 1. Effect of the GT box mutation on the Vα11.1 promoter and deletional analysis of the Vα2C promoter[a]

Construct	CAT activity		
	Expt 1		Expt 2 (%)
	%[b]	Fold[c]	
pCATMoEnT	0.57	1	
pTCR1700	0.26	0.5	
pTCR150MoEnT	0.26	0.5	
pTCR250MoEnT	7.93	14	6.56
Mut-1	0.10	0.2	0.11
Mut-2	0.06	0.1	0.07
pVα2C240MoEnT	5.40	9	
pVα2C470MoEnT	4.99	9	6.07
J21MoEn	23.27	40	45.45

[a] The constructs indicated were transfected into the Jurkat T-cell line, and CAT activities of the transfected cells were measured 2 days later. pCAT MoEnT contains the Moloney virus enhancer located at the 3′ end of the CAT gene without a promoter. Plasmids pTCR150MoEnT and pTCR250MoEnT contain various lengths of the Vα11.1 promoter region inserted upstream of the CAT gene in plasmid pCATMoEnT (see Fig. 1 and Materials and Methods); plasmids pVα2C240MoEnT and pVα2C470MoEnT contain the Vα2C promoter (see Materials and Methods). Mut-1 and Mut-2 are independently derived constructs of plasmid pTCR250MoEnT with a 3-bp GT box mutation. J21MoEn is a control construct containing a minimal c-*fos* promoter fragment and the Moloney virus enhancer in a CAT-containing plasmid.

[b] Percent acetylation of [^{14}C]chloramphenicol following incubation with the indicated protein extracts.

[c] Fold induction over pCATMoEnT background CAT activity.

domain. Therefore, the human Sp1 Zn finger domain was used as a probe to screen a human HUT78 (αβ T cells) cDNA library under low-stringency conditions. More than 20 positive clones were found, but most of these were Sp1 since they retained the hybridized signal after repeated stringent washes. Two of the clones (S1 and S13), however, lost their hybridization signal to the Sp1 Zn finger after stringent washing, indicating that they are homologous but not identical to Sp1. These two clones were chosen for further analysis as putative cDNAs encoding GT box-binding proteins. The original S1 clone was not full length, as revealed by its inability to produce in vitro-translated protein and by its small insert size. A full-length version of S1 (S1/7) was then isolated from a human Molt13 (γδ T cells) cDNA library and characterized. Sequencing analysis of the two clones revealed that S1/7 and S13 are indeed two separate genes which constitute a novel Sp1 gene family (Fig. 6 and 7). We propose the names Sp2 and Sp3 for these two genes to reflect this fact.

The Sp2 cDNA contains an open reading frame of 495 amino acids (Fig. 6). The translation initiation site was assigned on the basis of homology to Kozak sequences and on the assumption that it starts with a methionine. The assignment, however, is still tentative, as the reading frame is open to the extreme 5′ end of the cDNA clone for another 118 amino acids and other nonmethionine codons can be used in some rare occasions. The Sp3 cDNA contains an open reading frame of 713 amino acids (Fig. 7). Assignment of the translation initiation site for Sp3 is particularly difficult because no AUG codon can be found in the first 218 amino acids of the reading frame. Assignment of AUG as the start codon predicts a short polypeptide of 494 amino acids. By using the in vitro-translated product, the size of Sp3 was estimated at close to 100 kDa (see below). Furthermore, extensive homology to Sp1 can be detected in the first 218 amino acids of the open reading frame preceding the AUG codon; hence, it is very unlikely that Sp3 translation starts at the AUG at nucleotide 653. Several non-AUG start sites have previously been identified for mammalian genes (21, 39). Sp3 translation could initiate at AUU at nucleotide 183 and GUG at nucleotide 291. The translation initiation site of Sp3 was tentatively assigned to AUU starting at nucleotide 183, for a predicted polypeptide of 653 amino acids.

The protein structures of Sp2 and Sp3 are very similar to that of Sp1. All of them contain three Zn fingers with the structure Cys_2-His_2. Homology to Sp1 at the Zn finger domain is 72% for Sp2 and 90% for Sp3 (Fig. 8). The predicted Sp3 protein has extensive homology to Sp1

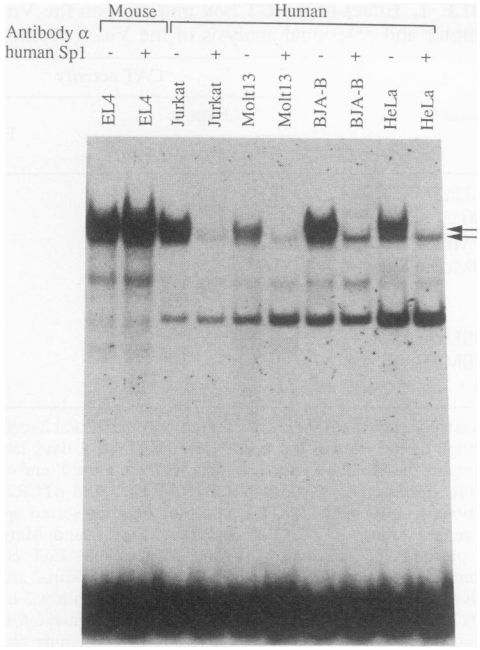

FIG. 5. Evidence that one of the GT box-binding proteins is transcription factor Sp1. Gel shift analysis using a GT box oligonucleotide was performed in the absence or presence of the mouse monoclonal antibody against human Sp1 (at region C immediately outside the Zn finger). As sources of the GT box-binding proteins, nuclear extracts from EL4 cells (mouse T-cell thymoma; used as a control, Jurkat and Molt13 cells (human T cells), BJA-B cells (human B cells), and nonlymphoid HeLa cells were used.

throughout the entire open reading frame, whereas homology outside the Zn finger domain for Sp2 is much more limited. Interestingly, both Sp2 and Sp3 also contain transactivation domains with characteristics similar to those of the Sp1 domains (Fig. 9). The Sp3 protein also shares with Sp1 the glutamine-rich domains A and B, while the Sp2 protein contains only one glutamine-rich region. All three proteins also contain a highly charged protein domain, C, located immediately 5′ of the zinc fingers and a serine/threonine-rich region. The Sp3 but not Sp2 protein contains a sequence homologous to the domain D sequence, which is located at the C-terminal end of the Sp1 protein. Hence, the Sp2 protein has a serine/threonine-rich region followed by a glutamine-rich domain (domain B), a highly charged domain (domain C), and three Zn fingers. The structure of the Sp3 protein, on the other hand, is strikingly similar to that of Sp1, with two homologous glutamine-rich regions, a serine/threonine-rich region, a highly charged domain, three zinc fingers, and a C-terminal domain similar to the Sp1 domain D (Fig. 9).

To compare the mRNA tissue distributions of Sp2 and Sp3 with that of Sp1, Northern blot analysis was performed. We found that Sp2 and Sp3 are widely expressed. The Sp2 message at 3.2 kb was found in the Jurkat T-cell line, the BJA-B B-cell line, K562 erythroid cells, and HeLa cells (Fig. 10 and data not shown). The Sp3 message of 4.2 kb was found in all cell lines tested, namely, mature T cells (Jurkat, Peer, and AN4.4), immature T cells (2052C and 2052D), mature and immature B cells (BJA-B and 300-19), erythroid cells (K562 and MEL), and nonlymphoid cells (HeLa and 3T3) (Fig. 9). Hence, all three members of the Sp1 family, Sp1, Sp2, and Sp3, are similar in their DNA-binding domains and transactivation domains as well as in their ubiquitous expression.

```
GCGGCCGCCGTAATGAGCGATCCACAGACC AGCATGGCTGCCACTGCTGCTGTGAGTCCC AGTGACTACCTGCAGCCTGCCGCCTCCACC    90
ACCCAGGACTCCCAGCCATCTCCCTTAGCC CTGCTTGCTGCAACATGTAGCAAAATTGGC CCTCCAGCAGTTGAAGCTGCTGTGACACCT   180
CCTGCTCCCCCACAGCCCACACCGCGGAAA CTTGTCCCTATCAAACCTGCCCCTCTCCCT CTCAGCCCCGGCAAGAATAGCTTTGGAATC   270
TTGTCCTCCAAAGGAAATATACTTCAGATT CAGGGGTCACAACTGAGCGCCTCCTATCCT GGAGGGCAGCTGGTGTTCGCTATCCAGAAT   360
CCCACCATGATCAACAAAGGGACCCGATCA AATGCCAATATCCAGTACCAGGCGGTCCCT CAGATTCAGGCAAGCAATTCCCAAACCATC   450
      M  I  N  K  G  T  R  S   N  A  N  I  Q  Y  Q  A  V  P   Q  I  Q  A  S  N  S  Q  T  I>
CAAGTACAGCCCAATCTCACCAACCAGATC CAGATCATCCCTGGCACCAACCAAGCCATC ATCACCCCCTCACCGTCCAGTCACAAGCCT   540
 Q  V  Q  P  N  L  T  N  Q  I   Q  I  I  P  G  T  N  Q  A  I   I  T  P  S  P  S  S  H  K  P>
GTCCCCATCAAGCCAGCCCCCATCCAGAAG TCGAGTACGACCACCACCCCCGTGCAGAGC GGGGCCAATGTGGTGAAGCTGACAGGTGGG   630
 V  P  I  K  P  A  P  I  Q  K   S  S  T  T  T  T  P  V  Q  S   G  A  N  V  V  K  L  T  G  G>
GGCGGCAATGTGACGCTCACTCTGCCCGTC AACAACCTTGTGAACGCCAGTGACACCGGG GCCCCTACTCAGCTCCTCACTGAAAGCCCC   720
 G  G  N  V  T  L  T  L  P  V   N  N  L  V  N  A  S  D  T  G   A  P  T  Q  L  L  T  E  S  P>
CCAACCCCGCTGTCTAAGACTAACAAGAAA GCAAGGAAGAAGAGCCTTCCTGCCTCCCAG CCCCCTGTGGCTGTGGCTGAGCAGGTGGAG   810
 P  T  P  L  S  K  T  N  K  K   A  R  K  K  S  L  P  A  S  Q   P  P  V  A  V  A  E  Q  V  E>
ACGGTGCTGATCGAGACCACCGCGGACAAC ATCATCCAGGCAGGAAATAACCTGCTCATT GTTCAGAGCCCTGGTGGGGGCCAGCCAGCT   900
 T  V  L  I  E  T  T  A  D  N   I  I  Q  A  G  N  N  L  L  I   V  Q  S  P  G  G  G  Q  P  A>
GTGGTCCAGCAGGTCCAGGTGGTGCCCCCC AAGGCCGAGCAGCAGCAGGTGGTACAGATC CCCCAGCAGGCTCTGCGGGTGGTGCAGGCG   990
 V  V  Q  Q  V  Q  V  V  P  P   K  A  E  Q  Q  Q  V  V  Q  I   P  Q  Q  A  L  R  V  V  Q  A>
GCATCTGCCACCCTCCCCACTGTACCCCAG AAGCCCTCCCAGAACTTTCAGATCCAGGCA GCTGAGCCGACACCTACTCAGGTCTACATC  1080
 A  S  A  T  L  P  T  V  P  Q   K  P  S  Q  N  F  Q  I  Q  A   A  E  P  T  P  T  Q  V  Y  I>
CGCACGCCTTCCGGTGAGGTGCAGACAGTC CTTGTCCAGGACAGCCCCCCAGCAACAGCT GCAGCCACCTCTAACACCACCTGTAGCAGC  1170
 R  T  P  S  G  E  V  Q  T  V   L  V  Q  D  S  P  P  A  T  A   A  A  T  S  N  T  T  C  S  S>
CCTGCATCCCGTGCTCCCCATCTGAGTGGG ACCAGCAAAAAGCACTCAGCTGCAATTCTC CGAAAAGAGCGTCCCCTGCCAAAGATTGCC  1260
 P  A  S  R  A  P  H  L  S  G   T  S  K  K  H  S  A  A  I  L   R  K  E  R  P  L  P  K  I  A>
CCAGCCGGGAGCATCATCAGCCTGAATGCA GCCCAGTTGGCGGCAGCTGCCCAGGCAATG CAGACCATCAACATCAATGGTGTCCAGGTC  1350
 P  A  G  S  I  I  S  L  N  A   A  Q  L  A  A  A  A  Q  A  M   Q  T  I  N  I  N  G  V  Q  V>
CAGGGCGTGCCTGTCACCATCACCAACACA GGCGGGCAGCAGCAGCTGACAGTGCAGAAT GTTTCTGGGAACAACCTGACCATCAGTGGG  1440
 Q  G  V  P  V  T  I  T  N  T   G  G  Q  Q  Q  L  T  V  Q  N   V  S  G  N  N  L  T  I  S  G>
CTGAGCCCCACCCAGATCCAGCTGCAAATG GAACAAGCCCTGGCCGGAGAGACCCAGCCC GGGGAGAAGCGGCGCCGCATGGCCTGCACG  1530
 L  S  P  T  Q  I  Q  L  Q  M   E  Q  A  L  A  G  E  T  Q  P   G  E  K  R  R  R  M  A  C  T>
TGTCCCAACTGCAAGGATGGGGAGAAGAGG TCTGGAGAGCAGGGCAAGAAGAAGCACGTT TGCCACATCCCCGACTGTGGCAAGACGTTC  1620
 C  P  N  C  K  D  G  E  K  R   S  G  E  Q  G  K  K  K  H  V   C  H  I  P  D  C  G  K  T  F>
CGTAAGACGTCCTTGCTGCGTGCCCATGTG CGCCTGCACACTGGCGAGCGGCCCTTTGTC TGCAACTGGTTCTTCTGTGGGAAGAGGTTC  1710
 R  K  T  S  L  L  R  A  H  V   R  L  H  T  G  E  R  P  F  V   C  N  W  F  F  C  G  K  R  F>
ACACGGAGTGACGAGCTCCAACGGCATGCT CGCACCCACACAGGGGACAAACGCTTCGAG TGCGCCCAGTGTCAGAAGCGCTTCATGAGG  1800
 T  R  S  D  E  L  Q  R  H  A   R  T  H  T  G  D  K  R  F  E   C  A  Q  C  Q  K  R  F  M  R>
AGTGACCACCTCACCAAGCATTACAAGACC CACCTGGTCACGAAGAACTTGTAAGGCCAA CTGCGGCGGGAGGCCTGAAGATGCAGTCCC  1890
 S  D  H  L  T  K  H  Y  K  T   H  L  V  T  K  N  L  *
CCACCTGTGTCCTCCCTGGGCCCCTGGTGG AAAGGAGCCCTGTGGCTGCCTTGGCCTGCC CTCAGCCCCACTCCTGTTCTGCAACTGTCC  1980
CCACAGGAAGGGGCTCTGTTCCCTGTATTG TCCTCCTTCTGAAGCCCCTTGTCTGCCTTG GCCCTTCCCCTCACCACGAGCTC         2063
```

FIG. 6. DNA and predicted amino acid sequences of the S1/7 cDNA clone (Sp2).

```
GAATTCCGGGCCATCGCCGGGCGACGACGA GGAGGAGGCGGCCGCCGCAGCCGGGGCCCC CGCCGCCGCCGGAGCGACAGGTGATTTGGC     90
TTCTGCACAGTTAGGAGGAGCACCAAACCG ATGGGAGGTTTTGTCAGCCACACCTACAAC TATAAAAGATGAAGCTGGTAATCTAGTCCA    180
GATTCCAAGTGCTGCTACTTCAAGTGGGCA GTATGTTCTTCCCCTTCAGAATTTGCAGAA TCAACAAATATTTTCCGTTGCACCAGGATC    270
  I  P  S  A  A  T  S  S  G  Q   Y  V  L  P  L  Q  N  L  Q  N   Q  Q  I  F  S  V  A  P  G  S>
AGATTCATCAAATGGTACAGTGTCCAGTGT TCAATATCAAGTGATACCACAGATCCAGTC AGCAGATGGTCAGCAGGTTCAAATTGGTTT    360
  D  S  S  N  G  T  V  S  S  V   Q  Y  Q  V  I  P  Q  I  Q  S   A  D  G  Q  Q  V  Q  I  G  F>
CACAGGCTCTTCAGATAATGGGGGTATAAA TCAAGAAAGCAGTCAAATTCAGATCATTCC TGGCTCTAATCAAACCTTACTTGCCTCTGG    450
  T  G  S  S  D  N  G  G  I  N   Q  E  S  S  Q  I  Q  I  I  P   G  S  N  Q  T  L  L  A  S  G>
AACACCTTCTGCTAACATCCAGAATCTCAT ACCACAGACTGGTCAAGTCCAGGTTCAGGG AGTTGCAATTGGTGGTTCATCTTTTCCTGG    540
  T  P  S  A  N  I  Q  N  L  I   P  Q  T  G  Q  V  Q  V  Q  G   V  A  I  G  G  S  S  F  P  G>
TCAAACCCAAGTAGTTGCTAATGTGCCTCT TGGTCTGCCAGGAAATATTACGTTTGTACC AATCAATAGTGTCGATCTAGATTCTTTGGG    630
  Q  T  Q  V  V  A  N  V  P  L   G  L  P  G  N  I  T  F  V  P   I  N  S  V  D  L  D  S  L  G>
ACTCTCGGGCAGTTCTCAGACAATGACTGC AGGCATTAATGCCGACGGACATTTGATAAA CACAGGACAAGCTATGGATAGTTCAGACAA    720
  L  S  G  S  S  Q  T  M  T  A   G  I  N  A  D  G  H  L  I  N   T  G  Q  A  M  D  S  S  D  N>
TTCAGAAAGGACTGGTGAGCGGGTTTCTCC TGATATTAATGAAACTAATACTGATACAGA TTTATTTGTGCCAACATCCTCTTCATCACA    810
  S  E  R  T  G  E  R  V  S  P   D  I  N  E  T  N  T  D  T  D   L  F  V  P  T  S  S  S  S  Q>
GTTGCCTGTTACGATAGATAGTACAGGTAT ATTACAACAAAACACAAATAGCTTGACTAC ATCTAGTGGGCAGGTTCATTCTTCAGATCT    900
  L  P  V  T  I  D  S  T  G  I   L  Q  Q  N  T  N  S  L  T  T   S  S  G  Q  V  H  S  S  D  L>
TCAGGGAAATTATATCCAGTCGCCTGTTTC TGAAGAGACACAGGCACAGAATATTCAGGT TTCTACAGCACAGCCTGTTGTACAGCATCT    990
  Q  G  N  Y  I  Q  S  P  V  S   E  E  T  Q  A  Q  N  I  Q  V   S  T  A  Q  P  V  V  Q  H  L>
ACAACTTCAAGAGTCTCAGCAGCCAACCAG TCAAGCCCAAATTGTGCAAGGTATTACACC ACAGACAATCCATGGTGTGCAAGCCAGTGG   1080
  Q  L  Q  E  S  Q  Q  P  T  S   Q  A  Q  I  V  Q  G  I  T  P   Q  T  I  H  G  V  Q  A  S  G>
TCAAAATATATCACAACAGGCTTTGCAAAA TCTTCAGTTGCAGCTGAATCCTGGAACCTT TTTAATTCAGGCACAGACAGTGACCCCTTC   1170
  Q  N  I  S  Q  Q  A  L  Q  N   L  Q  L  Q  L  N  P  G  T  F   L  I  Q  A  Q  T  V  T  P  S>
TGGACAGGTAACTTGGCAAACGTTTCAAGT ACAAGGGGTCCAGAACTTGCAGAATTTGCA AATACAGAATACTGCTGCCCAACAAATAAC   1260
  G  Q  V  T  W  Q  T  F  Q  V   Q  G  V  Q  N  L  Q  N  L  Q   I  Q  N  T  A  A  Q  Q  I  T>
TTTGACGCCTGTTCAAACCCTCACACTTGG TCAAGTTGCGGCAGGTGGAGCCTTCACTTC AACTCCAGTTAGTCTAAGCACTGGTCAGTT   1350
  L  T  P  V  Q  T  L  T  L  G   Q  V  A  A  G  G  A  F  T  S   T  P  V  S  L  S  T  G  Q  L>
GCCAAATCTACAAACAGTTACAGTGAACTC TATAGATTCTGCTGGTATACAGCTACATCC AGGAGAGAATGCTGACAGTCCTGCAGATAT   1440
  P  N  L  Q  T  V  T  V  N  S   I  D  S  A  G  I  Q  L  H  P   G  E  N  A  D  S  P  A  D  I>
TAGGATCAAGGAAGAAGAACCTGATCCTGA AGAGTGGCAGCTCAGTGGTGATTCTACCTT GAATACCAATGACCTAACACACTTAAGAGT   1530
  R  I  K  E  E  E  P  D  P  E   E  W  Q  L  S  G  D  S  T  L   N  T  N  D  L  T  H  L  R  V>
ACAGGTGGTAGATGAAGAAGGGGACCAACA ACATCAAGAAGGAAAAAGACTTCGGAGGGT AGCTTGCACCTGTCCCAACTGTAAAGAAGG   1620
  Q  V  V  D  E  E  G  D  Q  Q   H  Q  E  G  K  R  L  R  R  V   A  C  T  C  P  N  C  K  E  G>
TGGTGGAAGAGGTACCAATCTTGGGAAAAA GAAGCAACACATTTGTCATATACCAGGATG TGGTAAAGTCTATGGGAAGACCTCACATCT   1710
  G  G  R  G  T  N  L  G  K  K   K  Q  H  I  C  H  I  P  G  C   G  K  V  Y  G  K  T  S  H  L>
GAGAGCTCATCTGCGTTGGCATTCTGGAGA ACGCCCTTTTGTTTGTAACTGGATGTACTG TGGTAAAAGATTTACTCGAAGTGATGAATT   1800
  R  A  H  L  R  W  H  S  G  E   R  P  F  V  C  N  W  M  Y  C   G  K  R  F  T  R  S  D  E  L>
ACAGAGGCACAGAAGAACACATACAGGTGA GAAGAAATTTGTTTGTCCAGAATGTTCAAA ACGCTTTATGAGAAGTGACCACCTTGCCAA   1890
  Q  R  H  R  R  T  H  T  G  E   K  K  F  V  C  P  E  C  S  K   R  F  M  R  S  D  H  L  A  K>
ACATATTAAAACACACCAGAATAAAAAAGG TATTCACTCTAGCAGTACAGTGCTGGCATC TGTGGAAGCTGCGCGAGATGATACTTTGAT   1980
  H  I  K  T  H  Q  N  K  K  G   I  H  S  S  S  T  V  L  A  S   V  E  A  A  R  D  D  T  L  I>
TACTGCAGGAGGAACAACGCTTATCCTTGC AAATATTCAACAAGGTTCTGTTTCAGGGAT AGGAACTGTTAATACTTCCGCCACCAGCAA   2070
  T  A  G  G  T  T  L  I  L  A   N  I  Q  Q  G  S  V  S  G  I   G  T  V  N  T  S  A  T  S  N>
TCAAGATATCCTTACCAACACTGAAATACC TTTACAGCTTGTCACAGTTTCTGGAAATGA GACAATGGAGTAAATATTACACAAATACTT   2160
  Q  D  I  L  T  N  T  E  I  P   L  Q  L  V  T  V  S  G  N  E   T  M  E  *
ATTCATTGTGGTTATTTTTATACAGTAGTG AGAAGAATATTGTTCCTAAGTTCTCCGGAA TTC                              2223
```

FIG. 7. DNA and predicted amino acid sequences of the S13 cDNA clone (Sp3).

Sp2 and Sp3 encode GT box-binding proteins. To investigate the DNA-binding activities of the Sp2 and Sp3 proteins, both S1/7 and S13 cDNAs were subcloned into an pSP72 vector (Promega) which contains SP6 and T7 RNA promoters for in vitro transcription. In vitro transcription and translation of both the S1/7 and S13 clones were first performed in the presence of [^{35}S]methionine to test whether Sp2 and Sp3 proteins could be translated in vitro without an artificial AUG start codon. Consistent with the notion that both S1/7 and S13 clones are full length, the in vitro-translated products yielded an Sp2 protein of approximately 80 kDa and an Sp3 protein of approximately 100 kDa (Fig. 11a). Smaller species also found in the gel are presumably minor degradation products of the full-length proteins.

To test the DNA-binding activities of Sp2 and Sp3 proteins, in vitro translation reactions in the absence of any radioactivity were performed, and the products were used in a gel shift experiment using either the 88-bp Vα11.1 promoter fragment or the small GT box oligonucleotide as a probe. As controls, we included Sp1 protein translated in vitro and the rabbit reticulocyte lysate alone, without any exogenous RNA added (Fig. 11b, lanes 1 and 4). Sp2 protein binds to the Vα11.1 promoter fragment, although weakly (lane 3). The Sp3 protein, however, binds with affinity comparable to that of Sp1 to both the Vα11.1 promoter fragment (lane 3) and the GT box oligonucleotide (lanes 5 to 9). The binding is specific, as revealed by competition studies using specific and nonspecific competitors (lanes 5 to 9). As expected from differences in predicted amino acid sequence between Sp1 and Sp3, the Sp3 protein binding activities cannot be abolished or supershifted with a mono-

```
Sp1  TATQLSQGANGWQIISSSSGATPTSKEQSGSSTNGSNGSESSKNRTVSGGQYVVAAAPNL   66
Sp3                                                IPSAAT.S....L-PLQ..   18

Sp1  QNQQVLTGLP------GVMPNIQYQVIPQFQTVDGQQLQ--FAATGAQVQQDGSGQIQII  118
Sp3  ....IFSVA.GSDSSN.TVSSV.......I.SA....V.IG.TGSSDNGGNQE.S.....   78

Sp1  PGANQQIITNRGSGGNIIAAMPNLLQQAVPLQ-GLANNVLSGQTQYVTNVPVALNGNITL  177
Sp2   MINKGTRSNANIQYQAVPQIQASNS.TIQV.PN.T.QIQIIP GTNQAIITPSPSSHKP   58
Sp3  ..S..TLLASGTPSA..QNLI.QTG.VQ.QGV-AIGGSSFP....V.A...LG.P....F  138

Sp1  LPVNSV--SAATLTPSSQAVT--ISSSGSQ-ESGSQPVTSGTTISSASLVSSQASSSSFF  232
Sp2  V.IKPA--PIQKSSTTTTP.Q--SGANVVKLTG.GGN..LTLPVNNLVNA.DTGAPTQLL  114
Sp3  V.I...DLDSLG.SG...TM.AG.NAD.HL-INTG.AMD.SDNSERTGERV.PDINET-N  196

Sp1  TNANSYSTTTTTSNMGIMNFTTSGSSGTNSQGQTPQRVSGLQGSDALNIQQNQTSGGSLQ  292
Sp2  .ESPPTPLSK.NKKARKKSLPA.QPPVAVAE-.VETVLIETTADNIIQAGN.LLIVQ.-P  172
Sp3  .DTDLFVP.SSS.QLPV-TIDST.ILQQ.TNSL.T-SSGQVHS..-.QGNYI.SPVSEET  253

Sp1  AGQQKEGEQNQQTQQQQILIQPQLVQG-GQALQALQAAPLSGQTFTTQAISQETLQNLQL  351
Sp2  G.G.PAVV.QV.VVPPKA-E.Q.V..I-P.--...RVVQAASA.LP.--VP.KPS..F.I  226
Sp3  QA.NIQVSTA.PVV.HLQ.QES.QPTSQA.IV.GITPQTIH.VQASG.N...QA......  313

Sp1  QAV-PNSGPIIIRTPTVGPNGQVSWQTLQLQNLQV-QNPQAQTITLAPMQGVSLGQTSSS  409
Sp2  ..AE.TPTQVY....SGEVQTVLV-.DSP-PATAA-ATSNT-.CSSPASRAPH.SG..KK  282
Sp3  .-L-.P.T-FL.QAQ..T.S...T...F.V.GV.NL..L.I.-N.A.Q-.IT-.--.P--  363

Sp1  NTTLTPIASAASIPAGTVTVNAAQLSSMPGLQTINLSALGTSGIQVHPIQGL--PLAIAN  467
Sp2  -HSA-A.LRKERPLPKIAPAGSIISLNAAQ.AA-AAQ.MQ.IN.NGVQV..VPVTITNTG  339
Sp3  VQ...LGQV..GGAF-.S.PVSLSTGQL.N...VTVNSIDSA...L..GENADS.AD.RI  422

Sp1  APGDHGA-QLGLHGAGGDGIHDDTAGGEEGENSPDAQPQAGRRTRREACTCPYCKDSEGR  526
Sp2  GQQQLTV-.NVSGNNLTISGLSP.Q-IQLQM.QALAGET.P.EKR..M.....N...G.K  397
Sp3  KEEEPEDPEWQ.S.DSTLNTN.L.HLRVQVVDEEG.Q.H.E.K.L..V.....N..EGG.  482

Sp1  GSGDPGKKKQHJICHIOGCGKVYGKTSHLRAHLRWHTGERPFMCTWSYCGKRFTRSDELQ  586
Sp2  .-..EQ....-.V...PD...TFR...L....V.L.....V.N.FF.............  455
Sp3  ..T-NL.....P.....................S.....V.N.M.............  541

Sp1  RHKRTHTGEKKFACPECPKRFMRSDHLSKHIKTHQNKKGGPGVALSVGTLPLDSGAGSEG  646
Sp2  ..A....D.R.E.AQ.Q.........T..Y...LVT.NL*                    495
Sp3  ..R.....V.....S.........A.....IHSSSTVLASVEAARDDTLIT         601

Sp1  SGTATPSALITTNMVAMEAICPEGIARLANSGINVMQVADLQSINISGNGF*          697
SP3  A.GT.LI-.ANIQQGSVSG.GTVNTSATS.QD.LTNTEIP..LVTV...ETME*       653
```

FIG. 8. Amino acid comparison of Sp1, Sp2, and Sp3 proteins. Dots indicate identity. The Zn fingers for Sp1, Sp2, and Sp3 are underlined.

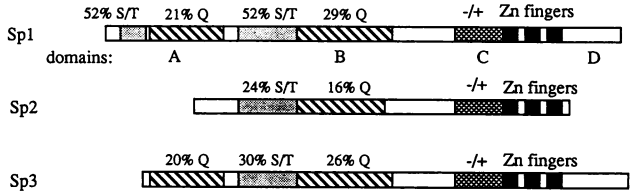

FIG. 9. Schematic diagram of predicted Sp2 and Sp3 protein domains. Transactivation domains similar to those in Sp1 can be found in both Sp2 and Sp3. Boxes indicate protein domains. Glutamine (Q) and serine/threonine (S/T) contents in each region are indicated. The transactivation domains for Sp1 (3, 4, 18, 19) are indicated as regions A through D. Similar regions B and C are found in Sp2, whereas Sp3 contain homologous regions A, B, C, and D. For Sp2, domain B (positions 145 to 264) contains 20 glutamine out of 120 amino acids (16% Q rich). For Sp3, domain A (positions 10 to 123) contains 23 glutamine out of 114 amino acids (20% Q rich) and domain B (positions 226 to 353) has 33 glutamine out of 128 amino acids (26% Q rich). Only one serine/threonine-rich region is found in Sp2 or Sp3. −/+, highly charged domain.

clonal antibody against Sp1 (lanes 10 to 13). Hence, both Sp2 and Sp3 proteins can bind to the GT box motif of the T-cell receptor Vα11.1 promoter region.

As Sp3 protein binds to the GT box of the TCR Vα11.1 promoter with high affinity, it might regulate the promoter activity. Indeed, a preliminary experiment indicated that Sp3 can transactivate the Vα11.1 GT box element in a *Drosophila* cell line (19a). This notion is further supported by the fact that the Sp3-GT box oligonucleotide complex comigrates with the second band of the GT box protein complex detected in the nuclear extracts (arrows in Fig. 12). Indeed, the purified Sp1-GT box complex migrates slower than the Sp3-GT box complex, and each complex comigrates with one of the bands detected in nuclear extracts. This occurs with use of either the GT box or the Sp1 consensus sequence, the GC box, as a probe (Fig. 12). Hence, other than a slight shift in migration pattern, no discernible differences in DNA-binding activity can be detected between Sp1 and the newly characterized Sp3 gene, whereas the Sp2 protein binds weakly to the TCR Vα GT box and not at all to the GC box (data not shown).

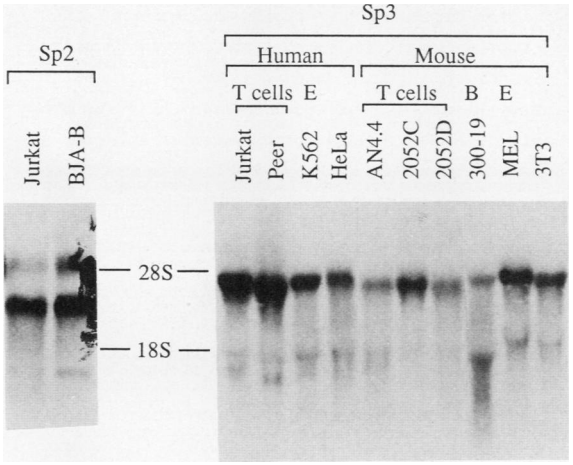

FIG. 10. Northern blot analysis of Sp2 and Sp3 cDNA clones. Northern blot analysis was performed with total RNAs from several human and mouse cell lines. Jurkat and Peer are human T leukemic cells, AN4.4 is a mouse T-cell hybridoma, K562 and MEL are erythroid cells of human and mouse origin, respectively, BJA-B is a human B-cell line, and 300-19 is a mouse B-cell line. RNAs from two mouse pre-T-cell lines (2052C and 2052D) as well as RNAs from nonlymphoid HeLa and 3T3 cells were also included. The probes used were a 1.5-kb *Eco*RI fragment of the S1 clone for Sp2 and a 2.4-kb *Eco*RI fragment of the S13 clone for Sp3.

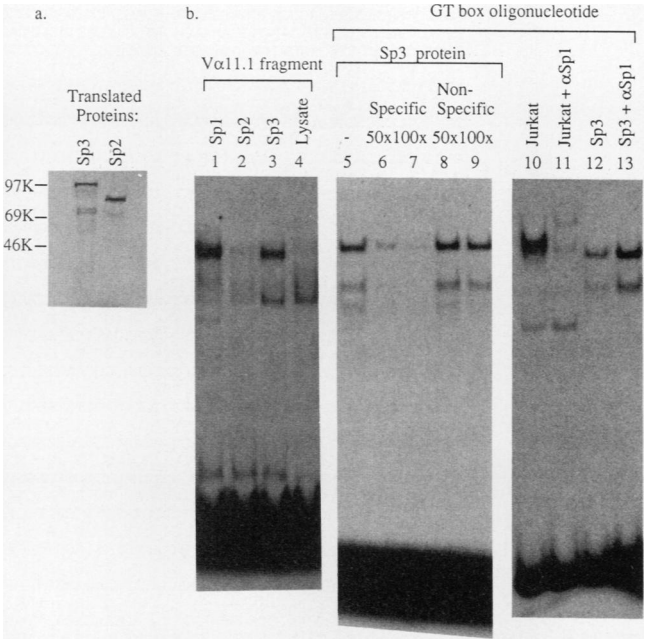

FIG. 11. Sp2 and Sp3 in vitro-translated products. (a) The in vitro-translated Sp2 and Sp3 products were labeled in the presence of [^{35}S]methionine. The resulting reaction products were run onto a sodium dodecyl sulfate–12% polyacrylamide gel. The Sp2 and Sp3 products were visualized by autoradiography. (b) Gel shift analysis was done with in vitro-translated proteins from Sp1, Sp2, and Sp3 RNAs generated from in vitro transcription of the respective cDNA clones. The in vitro translations were done without radioactivity, and gel shift analysis was done with 88 bp of the Vα11.1 fragment (lanes 1 to 4) or a small GT box oligonucleotide (lanes 5 to 13) as the probe. Specific and nonspecific DNAs were included in lanes 6 to 9 to test specificity of the Sp3 binding activity. Monoclonal antibody against human Sp1 was added for the binding reactions in lanes 11 and 13.

DISCUSSION

We have analyzed the T-cell-specific TCR Vα11.1 promoter and narrowed its activity to a 130-bp DNA sequence. Analysis of the minimal promoter region revealed a transcriptional element which we termed the GT box. The GT box is a crucial element for TCR α-promoter activity, and it binds to ubiquitously expressed proteins. The GT box is present in the TCR α enhancer as well as in three of four Vα upstream sequences analyzed. It is found in the upstream promoter regions of Vα11.1, Vα5H, and VαBDFL-1 but not Vα2C (Vα11.1 and Vα5H upstream regions were sequenced in this laboratory. VαBDFL-1 and Vα2C upstream sequences are the only two available in the GenBank). The absence of a GT box in a functional Vα 2C promoter indicates the presence of at least two classes of Vα gene segments, as judged from their transcriptional controlling elements.

The heterogeneity of the Vα promoter elements is in contrast to the homogeneity found in the TCR β and immu-

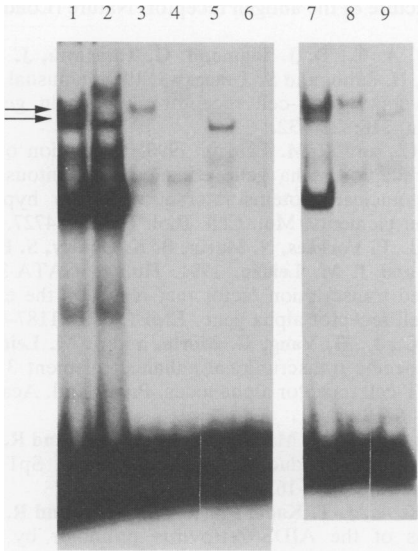

FIG. 12. Sp3 protein is the other protein binding to the GT box of the Vα11.1 promoter. Shown is gel shift analysis of HeLa nuclear extracts (lanes 1, 2, and 7), the in vitro-translated products of Sp1 (lanes 3, 4, and 8) and Sp3 (lanes 5 and 9), and rabbit reticulocyte lysate alone (lane 6). Monoclonal antibody against Sp1 was added for the binding reactions in lanes 2 and 4. The probes used were GC box (lanes 1 to 6) and GT box (lanes 7 to 9) oligonucleotides. Arrows indicate the protein-DNA complexes corresponding to Sp1 and Sp3. The Sp1 complex is supershifted with addition of monoclonal antibody against Sp1.

noglobulin heavy-chain gene promoters. All of the TCR β promoters contain a conserved and essential element homologous to the CRE (cyclic AMP-responsive element)/AP1 site (1, 2), and all of the immunoglobulin heavy-chain promoters contain the Oct element (reference 33 and references therein). This diversity in the Vα gene families might reflect the unique organization of the TCR α/δ locus (27). The δ locus is located within the TCR α-gene segments, and rearrangement as well as transcription of different α and δ gene segments are highly regulated. In T cells committed to the αβ lineage, the α-specific V-gene segments are rearranged to the respective D/Jδ gene segments. Some V-gene segments, however, can rearrange to either the α or δ locus. As each V-gene segment contains its own promoter at the 5′ upstream region, it is possible that different specificities of sterile transcription from these V-gene promoters could contribute to differences in accessibility to recombinases during T-cell development. This could result in distinct but overlapping gene segment usage in αβ and γδ T cells (35). We showed that the three Vα gene segments analyzed from the Vα1, -6, and -11 families contain GT box-binding sites and that one from the Vα3 family (Vα2C) does not have the GT box in its functional promoter region. Interestingly, members of the Vα1, -6, and -11 gene families have been found to rearrange to both α and δ genes, whereas members of the Vα3 family have thus far never been found in γδ T cells (17). Hence, sterile transcription of Vα1, -6, and -11 families from their respective GT box-containing promoters might occur in both αβ and γδ T cells, whereas a sterile transcript from the Vα3 GT-less promoter occurs only in αβ T cells. The result might be accessibility to recombinase and rearrangement of Vα1, -6, and -11 gene segments in both αβ and γδ T-cell lineages. The absence of germ line transcript from the Vα3 gene segment in the γδ lineage, on the other hand, might lead to its inaccessibility to recombinase in γδ T cells. More experiments (e.g., using transgenic mice), however, are needed to establish the functional relationship between the presence or absence of the GT box promoter element and the ability of different Vα gene segments to rearrange to the TCR α or δ gene.

In this study, we have characterized the GT box-binding proteins and shown that two proteins bind to the Vα promoter region. Using a monoclonal antibody specific for Sp1 protein, we showed that one of them is likely to be Sp1. To isolate additional GT box-binding proteins, we have used the Sp1 zinc finger domain as a probe and have isolated two novel Sp1-like cDNA clones which we termed Sp2 and Sp3 because of their extensive homology to Sp1 at the Zn finger region. Protein homology between Sp2 and Sp1 at the zinc finger domain is 74%, whereas Sp3 is closely related to Sp1, with homology at the Zn finger domain of 90%. Furthermore, the homology to Sp1 extends outside the Sp3 Zn finger domain. Previous studies have defined four transactivation domains in Sp1 as domains A and B, which are glutamine rich, domain C, which is highly charged, and domain D, which is located at the most carboxy-terminal portion of Sp1. The predicted Sp3 protein contains regions homologous to all four transactivation domains, with both of Sp3 domains A and B containing a glutamine level comparable to that in the Sp1 counterparts. Hence, Sp3 is closely related to Sp1 in many aspects of protein structure and presumably function as well. The predicted Sp2 protein, on the other hand, seems to contain fewer transactivation domains than does either Sp1 or Sp3. The predicted Sp2 protein has only a 16% glutamine-rich region, similar to the Sp1 B domain.

Sp3 is most likely the second Vα11.1 GT box-binding protein detected in nuclear extracts because in gel shift experiments, the DNA-Sp3 protein complex comigrates with the DNA-protein complex from nuclear extract. Sp3 mRNA is also expressed widely, and the in vitro-translated product binds to the GT box element of the TCR promoter with high affinity, consistent with characteristics of the GT box-binding proteins from nuclear extracts. Furthermore, preliminary experiments indicated that Sp3 can transactivate the Vα11.1 GT box in a *Drosophila* cell line (19a). The Sp2 protein, on the other hand, binds to the GT box in the TCR promoter very weakly. Because flanking sequences might affect binding affinity, Sp2 might bind to other GT box motifs which regulate other genes. More interestingly, Sp3 binds to the Sp1 consensus sequence GC box as well, although both Sp3 and Sp1 bind to the GT box with higher affinity. The striking similarity between Sp1 and Sp3 raises the interesting issue of whether an activity attributed to Sp1 in many previous studies might actually be due to the Sp3 protein. It also raises a more general issue of how gene regulation by two very similar factors is actually achieved in vivo.

The GT box is also present in the human and mouse TCR α enhancers (13, 36). There are two GT boxes in the human α enhancer at approximately 60 and 100 bp 3′ from the GATA3-binding site (12, 16, 20, 24) and one in the mouse enhancer 100 bp away from the GATA3-binding site. Although the GT box is located outside the original minimal enhancer fragment as defined by transient transfection experiments, it might reflect redundancy in enhancer function or it might have another function in TCR α gene regulation (e.g., serving as an LCR). Redundancy in the TCR α enhancer was demonstrated by experiments deleting one of the α enhancer protein-binding site (CRE = Tα1). Deletion of the Tα1 site in the minimal α enhancer resulted in

abrogation of the enhancer activity, whereas a similar deletion on a larger piece of the enhancer had a negligible effect (the larger enhancer fragment includes the GT box as well as an additional protein-binding site, Tα4 [11]). A possible function of the GT box as an LCR was previously suggested for the β-globin LCR (16; for a review of the globin system, see reference 25). The GT box (=AC box) is present in all four DNase I-hypersensitive sites which constitute the β-globin LCR. In addition, the GT box is also found in α- and ε-globin promoters and β-globin enhancer (see reference 26 and references therein). The minimal 225-bp fragment containing HSS2 of the β-globin LCR in particular (26) contains four GT box-binding sites as well as several GATA-binding sites. GATA- and GT-binding sites are common features of many erythroid cell-specific regulatory elements. It was proposed that interactions of erythroid-specific GATA1- and GT box-binding proteins are a key component of the LCR activity in erythroid cells (26). As similar GATA and GT box elements are also found in the TCR α enhancer, the interaction of T-cell-specific GATA3 protein and members of the Sp1 multigene family might play a key role in any possible T-cell-specific LCR function. Finally, GC and GT boxes are found in regulatory elements of other genes as well, including the human immunodeficiency virus long terminal repeat (15), SV40 enhancer (39), and BPV promoters (23, 32). Hence, this novel Sp1 multigene family might regulate a wide variety of viral and cellular genes in different tissues.

ACKNOWLEDGMENTS

We thank Bob Tjian for the gift of the Sp1 monoclonal antibody, G. Tanigawa and M. Steinmetz for providing genomic Vα clones, Joseph Marine, Virginia Smith, Kirk Johnson, and Nilabh Shastri for critical reading of the manuscript, and Yoon Choi for expert technical assistance.

A.W. is a Searle scholar, Cancer Research Institute investigator, and Leukemia Society of America special fellow. This work was also supported by the Markey Innovation Fund and NIH grant RO1 to A.W.

REFERENCES

1. **Anderson, S. J., H. S. Chou, and D. Y. Loh.** 1988. A conserved sequence in the T-cell receptor beta-chain promoter region. Proc. Natl. Acad. Sci. USA **85**:3551–3554.
2. **Anderson, S. J., S. Miyake, and D. Y. Loh.** 1989. Transcription from a murine T-cell receptor V beta promoter depends on a conserved decamer motif similar to the cyclic AMP response element. Mol. Cell. Biol. **9**:4835–4845.
3. **Courey, A. J., D. A. Holtzman, S. P. Jackson, and R. Tjian.** 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell **59**:827–836.
4. **Courey, A. J., and R. Tjian.** 1988. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell **55**:887–898.
5. **Davis, M. M., and P. J. Bjorkman.** 1988. T-cell antigen receptor genes and T-cell recognition. Nature (London) **334**:395–402.
6. **Dembic, Z., W. Haas, S. Weis, J. McCubrey, H. Kiefer, H. von Boehmer, and M. Steinmetz.** 1986. Transfer of specificity by murine alpha and beta T-cell receptor genes. Nature (London) **320**:232–238.
7. **Ferrier, P., B. Krippl, T. K. Blackwell, A. J. Furley, H. Suh, A. Winoto, W. D. Cook, L. Hood, F. Costantini, and F. W. Alt.** 1990. Separate elements control DJ and VDJ rearrangement in a recombination substrate. EMBO J. **9**:117–125.
8. **Ferrier, P., B. Krippl, A. J. Furley, T. K. Blackwell, H. Suh, M. Mendelsohn, A. Winoto, W. D. Cook, L. Hood, and F. Costantini.** 1989. Control of VDJ recombinase activity. Cold Spring Harbor Symp. Quant. Biol. **1**:191–202.
9. **Fink, P. J., L. A. Matis, D. L. McElligott, M. Bookman, and S. M. Hedrick.** 1986. Correlation between T cell specificity and the structure of the antigen receptor. Nature (London) **321**:219–226.
10. **Hayday, A. C., D. J. Diamond, G. Tanigawa, J. S. Heilig, V. Folsom, H. Saito, and S. Tonegawa.** 1985. Unusual organization and diversity of T-cell receptor alpha-chain genes. Nature (London) **316**:828–832.
11. **Ho, I. C., and J. M. Leiden.** 1990. Regulation of the human T-cell receptor alpha gene enhancer: ubiquitous and T-cell-specific nuclear proteins interact with four hypomethylated enhancer elements. Mol. Cell. Biol. **10**:4720–4727.
12. **Ho, I. C., P. Vorhees, N. Marin, B. K. Oakley, S. F. Tsai, S. H. Orkin, and J. M. Leiden.** 1991. Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor alpha gene. EMBO J. **10**:1187–1192.
13. **Ho, I. C., L. H. Yang, G. Morle, and J. M. Leiden.** 1989. A T-cell-specific transcriptional enhancer element 3′ of C alpha human T-cell receptor alpha locus. Proc. Natl. Acad. Sci. USA **86**:6714–6718.
14. **Jackson, S. P., J. J. MacDonald, M. S. Lees, and R. Tjian.** 1990. GC box binding induces phosphorylation of Sp1 by a DNA-kinase. Cell **63**:155–165.
15. **Jones, K. A., J. T. Kadonaga, P. A. Luciw, and R. Tjian.** 1986. Analysis of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. Science **232**:755–759.
16. **Joulin, V., D. Bories, J. F. Eleouet, M. C. Labastie, S. Chretien, M. G. Mattei, and P. H. Romeo.** 1991. A T-cell specific TCR delta DNA binding protein is a member of the human GATA family. EMBO J. **10**:1809–1816.
17. **Jouvin, M. E., I. Hue, P. N. Marche, G. C. Liebe, J. P. Marolleau, B. Malissen, P. A. Cazenave, and M. Malissen.** 1990. Genomic organization of the mouse T cell receptor V alpha family. EMBO J. **9**:2141–2150.
18. **Kadonaga, J. T., K. R. Carner, F. R. Masiarz, and R. Tjian.** 1987. Isolation of cDNA encoding transcription factor Sp1 and analysis of the DNA binding domain. Cell **51**:1079–1090.
19. **Kadonaga, J. T., A. J. Courey, J. Ladika, and R. Tjian.** 1988. Distinct regions of Sp1 modulate DNA binding and activation. Science **242**:1566–1570.

19a. **Kingsley, C., and A. Winoto.** Unpublished data.

20. **Ko, L. J., M. Yamamoto, M. W. Leonard, K. M. George, P. Ting, and J. D. Engel.** 1991. Murine and human T-lymphocyte GATA-3 factors mediate transcription through a cis-regulatory element within the human T-cell receptor delta gene enhancer. Mol. Cell. Biol. **11**:2778–2784.
21. **Kozak, M.** 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA **87**:8301–8305.
22. **Lenardo, M., J. W. Pierce, and D. Baltimore.** 1987. Protein-binding sites in Ig gene enhancers determine activity and inducibility. Science **236**:1573–1577.
23. **Li, R., J. D. Knight, S. P. Jackson, R. Tjian, and M. R. Botchan.** 1991. Direct interaction between Sp1 and the BPV enhancer E2 protein mediates synergistic activation of transcription. Cell **65**:493–505.
24. **Marine, J., and A. Winoto.** 1991. The human enhancer-binding protein Gata3 binds to several T-cell regulatory elements. Proc. Natl. Acad. Sci. USA **88**:7284–7288.
25. **Orkin, S. H.** 1990. Globin gene regulation and switching: circa 1990. Cell **63**:665–672.
26. **Philipsen, S., D. Talbot, P. Fraser, and F. Grosveld.** 1990. The beta-globin dominant control region: hypersensitive site 2. EMBO J. **9**:2159–2167.
27. **Raulet, D. H.** 1989. The structure, function, and molecular genetics of the gamma/delta T receptor. Annu. Rev. Immunol. **7**:175–207.
28. **Schlissel, M. S., and D. Baltimore.** 1989. Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. Cell **58**:1001–1007.
29. **Sen, R., and D. Baltimore.** 1986. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell **46**:705–716.
30. **Singh, H., R. Sen, D. Baltimore, and P. Sharp.** 1986. A nuclear factor that binds a conserved sequence motif in transcriptional

control elements of immunoglobulin genes. Nature (London) **319**:154–158.
31. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as a transcription control element. Cell **57**:103–113.
32. **Spalholz, B. A., S. B. Vande Pol, and P. M. Howley.** 1991. Characterization of the *cis* elements involved in basal and E2-transactivated expression of the bovine papillomavirus P_{2443} promoter. J. Virol. **65**:743–753.
33. **Staudt, L. M., R. G. Clerc, H. Singh, J. H. LeBowitz, P. A. Sharp, and D. Baltimore.** 1988. Cloning of a lymphoid-specific cDNA encoding a protein binding regulatory octamer DNA motif. Science **241**:577–580.
33a. **Winoto, A.** Unpublished data.
34. **Winoto, A., and D. Baltimore.** 1989. Alpha beta lineage-specific expression of the alpha T cell by nearby silencers. Cell **59**:649–655.
35. **Winoto, A., and D. Baltimore.** 1989. Developmental regulation of the TCR alpha delta locus. Cold Spring Harbor Symp. Quant. Biol. **1**:87–92.
36. **Winoto, A., and D. Baltimore.** 1989. A novel, inducible and T cell-specific enhancer located at the 3′ end of the T cell receptor alpha locus. EMBO J. **8**:729–733.
37. **Winoto, A., S. Mjolsness, and L. Hood.** 1985. Genomic organization of the genes encoding mouse T-cell receptor alpha-chain. Nature (London) **316**:832–836.
38. **Winoto, A., J. L. Urban, N. C. Lan, J. Goverman, L. Hood, and D. Hansburg.** 1986. Predominant use of a V alpha gene segment in mouse T-cell receptors for cytochrome c. Nature (London) **324**:679–682.
39. **Xiao, J. H., I. Davidson, H. Matthes, J. M. Garnier, and P. Chambon.** 1991. Cloning, expression, and transcriptional properties of the human factor TEF-1. Cell **65**:551–568.